

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 5 (bases 1 to 126803)  
 Waterston, R.  
 Direct Submission  
 Submitted (21-DEC-1999) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Jul 17, 1999 this sequence version replaced gi:4371313.

## COMMENT

----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: sapiens@wustl.wustl.edu  
 ----- Summary Statistics  
 Center project name: H\_NH0022N19

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 clone sections once, or longer because we provide a small overlap  
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by  
 restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and  
 sequencing collaboration between the NGR1 Chromosome 7 Mapping  
 Project (Eric D. Green, Director), John D. McPherson in the  
 Department of Genetics (Washington University), and the Washington  
 University Genome Sequencing Center. For additional information  
 about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPT-11 human BAC library was made from the blood of one male  
 donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Fengsen, E.,  
 Tateo, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
 and coworkers at the Roswell Park Cancer Institute  
 (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTA-332P12, 200 bp overlap. The  
 clone sequenced to the right is CTR-11H14, 200 bp overlap. Actual  
 start of this clone is at base position 81540 of CTA-332P12; actual  
 end is at base position 14809 of CTR-11H14.

## FEATURES

## source

Location/Qualifiers

1..126803

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7q42"

/clone="RP11-22N19"

/clone\_lib="RPT-11"

28..136

/rpt\_family="Retroviral"

repeat\_region

806..926

/rpt\_family="MER1\_type"

repeat\_region

1018..1961

/rpt\_family="L2"

repeat\_region

2105..2330

/rpt\_family="MIR"

repeat\_region

/rpt\_family="Alu"  
 3032..3102  
 repeat\_region  
 /rpt\_family="L2"  
 3625..3739  
 repeat\_region  
 /rpt\_family="L2"  
 3966..4242  
 repeat\_region  
 /rpt\_family="Alu"  
 4660..4731  
 repeat\_region  
 /rpt\_family="GA-rich"  
 5436..5721  
 repeat\_region  
 /rpt\_family="MIR"  
 5788..6025  
 repeat\_region  
 /rpt\_family="MIR"  
 6341..6622  
 repeat\_region  
 /rpt\_family="Alu"  
 6738..6759  
 repeat\_region  
 /rpt\_family="AT-rich"  
 7699..7909  
 repeat\_region  
 /rpt\_family="MIR"  
 8132..8256  
 repeat\_region  
 /rpt\_family="L1"  
 9057..10448  
 repeat\_region  
 /rpt\_family="MER2\_type"  
 10454..10584  
 repeat\_region  
 /rpt\_family="MER2\_type"  
 11748..11829  
 repeat\_region  
 /rpt\_family="MER116"  
 11833..12128  
 repeat\_region  
 /rpt\_family="Alu"  
 12401..12763  
 repeat\_region  
 /rpt\_family="Retroviral"  
 13157..13457  
 repeat\_region  
 /rpt\_family="Alu"  
 13822..13920  
 repeat\_region  
 /rpt\_family="U6"  
 13923..14059  
 repeat\_region  
 /rpt\_family="L1"  
 14275..14376  
 repeat\_region  
 /rpt\_family="MIR"  
 16260..17099  
 repeat\_region  
 /rpt\_family="L1"  
 17100..17462  
 repeat\_region  
 /rpt\_family="L1"  
 17449..21719  
 repeat\_region  
 /rpt\_family="L1"  
 21760..21844  
 repeat\_region  
 /rpt\_family="MIR"  
 22057..22460  
 repeat\_region  
 /rpt\_family="MALR"  
 22566..22899  
 repeat\_region  
 /rpt\_family="L2"  
 23023..23097  
 repeat\_region  
 /rpt\_family="L2"  
 23514..23578  
 repeat\_region  
 /rpt\_family="MIR"  
 23596..23883  
 repeat\_region  
 /rpt\_family="MER2\_type"  
 24118..24342  
 STS  
 /db\_xref="GI:454569"  
 25915..26030  
 repeat\_region  
 /rpt\_family="MIR"  
 26172..26631  
 repeat\_region  
 /rpt\_family="MALR"  
 27175..27456  
 repeat\_region  
 /rpt\_family="Alu"  
 27457..27504  
 repeat\_region  
 /rpt\_family="A-rich"  
 27506..27782  
 repeat\_region  
 /rpt\_family="Alu"  
 28153..28267  
 repeat\_region  
 /rpt\_family="MIR"  
 28272..28311  
 repeat\_region  
 /rpt\_family="(CTA)n"

```
repeat_region 29763. .30062
                /rpt_family="Alu"
repeat_region 30315. .30382
                /rpt_family="MaLR"
repeat_region 30494. .30563
                /rpt_family="MaLR"
repeat_region 30717. .30804
                /rpt_family="MER1_type"
repeat_region 31224. .31765
                /rpt_family="Other"
repeat_region 32610. .32885
                /rpt_family="Alu"
repeat_region 32995. .33038
                /rpt_family="T-rich"
repeat_region 33046. .34923
                /rpt_family="L1"
repeat_region 34924. .35226
                /rpt_family="Alu"
repeat_region 35227. .36291
                /rpt_family="L1"
repeat_region 36472. .36512
                /rpt_family="(TATG)n"
repeat_region 36514. .36800
                /rpt_family="Alu"
repeat_region 36977. .37133
```

```
Query Match      92.4%; Score 451.6; DB 9; Length 126803;
Best Local Similarity 98.9%; Pred. No. 3.7e-94;
Matches 454; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy      21 TGTACTATTATTGGCATATAGTTTCAAAAATTCACAGAAGGGAGCCAGGTGCTCTCATTG 80
          |||
Db      69732 TGTACTATTATTGGTATATAGTTTCAAAAATTCACAGAAGGGAGCCAGGTGCTCTCATTG 69673

Qy      81 CCTAACAAAATAATGGAAATATGTATTCATTCTAACATCTTGACATACAGTATAAAGGGC 140
          |||
Db      69672 CCTAACAAAATAATGGAAATATGTATTCATTCTAACATCTTGACATACAGTATAAAGGGC 69613

Qy     141 ACTCAGCAAGTGCTTTTAGTTAGACTGATTTTAAATGAGTAGATTACAGGAATACCAGAA 200
          |||
Db     69612 ACTCAGCAAGTGCTTTTAGTTAGACTGATTTTAAATGAGTAGATTACAGGAATACCAGAA 69553

Qy     201 AACGCACACACACAGATAATTGGTTGAAAATACGAGGTTAGAAACTTAACCATTTGAAG 260
          |||
Db     69552 AACGCACACACACAGATAATTGGTTGAAAATACGAGGTTAGAAACTTAACCATTTGAAG 69493

Qy     261 AGGAGTAAGGCTACTTAAAGCGTTAAAACTAATTGGGTAAGGTATGGTTGACCCAGCTA 320
          |||
Db     69492 AGGAGTAAGGCTACTTAAAGCGTTAAAACTAATTGGGTAAGGTATGGTTGACCCAGCTA 69433

Qy     321 CTTCAATTGCGCTTACGATGTATATTCATTAAACTAAGTCACTCACCTTCTCTGTGATGTT 380
          |||
Db     69432 CTTCAATTGCGCTTACGATGTATATTCATTAAACTAAGTCACTCACCTTCTCTGTGATGTT 69373

Qy     381 GATGGTTTGGTATAGTAAACATGAGATATGATTAAAGGTGATTACGGGATAGATCAAGTG 440
          |||
Db     69372 GATGGTTTGGTATAGTAAACATGAGATATGATTAAAGGTGATTACGGGATAGATCAAGTG 69313

Qy     441 TCTGCCTAAGTAAATCTGGGNTTCAATTTTTTTTCTAG 479
          |||
Db     69312 TCTGCCTAAGTAAATCTGTGTTTCATTTTTTTTCTAG 69274
```

RESULT 3  
 AC007032/c  
 LOCUS AC007032 126803 bp DNA linear PRI 21-DEC-1999  
 DEFINITION Homo sapiens BAC clone RP11-22N19 from 7q22, complete sequence.  
 ACCESSION AC007032  
 VERSION AC007032.2 GI:5523832  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 126803)  
 AUTHORS Sulston, J.E. and Waterston, R.H.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11): 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 126803)  
 AUTHORS Maas, J., Wohldmann, P., Harper, M. and Phillips, A.  
 TITLE The sequence of Homo sapiens BAC clone RP11-22N19  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 126803)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-MAR-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE 4 (bases 1 to 126803)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUL-1999) Genome Sequencing Center, Washington